

10/518377

WO 2004/003204

PCT/KR2003/001279

Sequence Listing

DT01 Rec'd PCT/PTC 28 DEC 2004

<110> Korea Research Institute of Bioscience and Biotechnology
LeadBio, Inc
Bio Holdings CO., LTD

<120> *Hansenula polymorpha* yapsin deficient mutant strain and process
for the preparation of recombinant proteins using the same

<160> 16

<170> KopatentIn 1.71

<210> 1

<211> 3151

<212> DNA

<213> *Hansenula polymorpha*

<220>

<221> sig_peptide

<222> (901)..(903)

<223> initiation codon

<220>

<221> 5'UTR

<222> (1)..(900)

<220>

<221> 3'UTR

<222> (2622)..(3151)

<220>

<221> terminator

<222> (2629)..(2631)

<223> termination codon

<220>

<221> CDS

<222> (901)..(2628)

<223> coding sequence

Sequence Listing

```

<400>    1
agttgagtgc caatagtgtg gcgaacttca aatgcctta ctgtccgcga acaaccacca      60

ttgcccaggc tgtgcaggcc agatttgtt aatttgtaaa agtgaaaaaa atttattccg      120

ctatgcctaa cccaaggagcc cgcaagaaga ggcggacaga agactttcc agctcttcgg      180

catctgaaaaa cgataagtgc tccgagagcg tgaccagtgt acaggaagag cagccggatg      240

cgcccgaaac atacacaata gatggcctgg acacgcaaga ggtgtctgac agcacacagg      300

tgagactcca acagctgaac gcagacaggt tggccagcat agagcaaagc ctttcaggca      360

acctcaaact ggacataaac gcagtacgcc agatagatga tgtgcgtgag cagctgcaga      420

acgagttttt gaagaaattt cttgtcacat attctgagga cctggatgcg ctgcgtcaga      480

aaaccgattt caaggaaaac tcactcaaaa ccctcgccccg tcttctcaaa gagagcggaa      540

acatatttga tcatggaaact ctcaagtgc tagttgagtg atgtatatga taatgtctaa      600

tttaattttt catcagtgtg caagatctgg gcttagccgt tctaaatggt atattcaggc      660

tgtgcaagcc acatttaaaa ttaccccatc ggttttaaa ttctattgtt agaaattagg      720

atctacatag agtagagtg agcaacagaa cattgtttgc tatccgggcc ctccgactgg      780

aacgtttac cttagctac tatttattca gaaaaaagag tgcattttca tctatcaagg      840

tctcaaagtg tcgaatcaaa tcactagtat ttttccgag actaaaaaaaaa agttgacaca      900

atg aaa gtt gct aca ctg ttt ttc ttg gct tcg agt gtc tgt gtg ctg      948
Met Lys Val Ala Thr Leu Phe Phe Leu Ala Ser Ser Val Cys Val Leu

    1          5          10          15

gga gac cca cag ttc gtg aaa ctg gag gcc tct gtt ctt cgg gga tcc      996
Gly Asp Pro Gln Phe Val Lys Leu Glu Ala Ser Val Leu Arg Gly Ser

    20         25         30

```

Sequence Listing

act tac aag gat tcc cag aag ggg gcc aag ccg ttc atg ttg gaa aag	1044		
Thr Tyr Lys Asp Ser Gln Lys Gly Ala Lys Pro Phe Met Leu Glu Lys			
35	40	45	
agg gct gat gac ggc tcg gtc acg atg gaa ttg cag aac gcc cag tct	1092		
Arg Ala Asp Asp Gly Ser Val Thr Met Glu Leu Gln Asn Ala Gln Ser			
50	55	60	
ttc tac caa gtc gag atc gag ata gga tct gat aag cag aag gtg ggg	1140		
Phe Tyr Gln Val Glu Ile Glu Ile Gly Ser Asp Lys Gln Lys Val Gly			
65	70	75	80
gtt ttg att gat acc ggt tcc tcg gac ttg tgg gtg atg aac tcg aat	1188		
Val Leu Ile Asp Thr Gly Ser Ser Asp Leu Trp Val Met Asn Ser Asn			
85	90	95	
aac tct tac tgt tcg tct tcc agc act aaa aaa ttg aaa cgg gac gga	1236		
Asn Ser Tyr Cys Ser Ser Ser Thr Lys Lys Leu Lys Arg Asp Gly			
100	105	110	
ccg gcc gat gcg cta caa aaa gga cgc gat ctt tcc gac ctg tac aat	1284		
Pro Ala Asp Ala Leu Gln Lys Gly Arg Asp Leu Ser Asp Leu Tyr Asn			
115	120	125	
ttc aac tct cca aac gaa gac aac aat gca aaa gga ttc ttg ggt ggc	1332		
Phe Asn Ser Pro Asn Glu Asp Asn Asn Ala Lys Gly Phe Leu Gly Gly			
130	135	140	
tgg gga gac ttg acc aca gta gag act gca acc cag gat gag aca cag	1380		
Trp Gly Asp Leu Thr Thr Val Glu Thr Ala Thr Gln Asp Glu Thr Gln			
145	150	155	160
acg gct ctc gct gcg cag gcc acc gtc gac tgc tcg cta tac gga acg	1428		
Thr Ala Leu Ala Ala Gln Ala Thr Val Asp Cys Ser Leu Tyr Gly Thr			
165	170	175	
ttc aat cct tca acg tcc aat tcg ttc cac aac aac ggc acc aca ttt	1476		
Phe Asn Pro Ser Thr Ser Asn Ser Phe His Asn Asn Gly Thr Thr Phe			
180	185	190	

Sequence Listing

gag att tcg tac gcg gac cgc act ttt gcc cgt gga acc tgg ggc tac 1524
 Glu Ile Ser Tyr Ala Asp Arg Thr Phe Ala Arg Gly Thr Trp Gly Tyr
 195 200 205

 gag gat gtc act ttc aat ggt gtc acg gtt aac gat ctc tcg ttg gcc 1572
 Asp Asp Val Thr Phe Asn Gly Val Thr Val Asn Asp Leu Ser Leu Ala
 210 215 220

 gtg gca gat gaa aca gat tct tcg act ggt gtt ttt ggt atc gga ttg 1620
 Val Ala Asp Glu Thr Asp Ser Ser Thr Gly Val Phe Gly Ile Gly Leu
 225 230 235 240

 agg gaa ttg gaa acc aca tac tca gga ggc gga cca cag cat tac atc 1668
 Arg Glu Leu Glu Thr Thr Tyr Ser Gly Gly Pro Gln His Tyr Ile
 245 250 255

 tac gac aac tta cct ttc aaa atg gtc gac cag gga ctc atc aat aga 1716
 Tyr Asp Asn Leu Pro Phe Lys Met Val Asp Gln Gly Leu Ile Asn Arg
 260 265 270

 gcc gcc tat tcc gtc tac ctg aac tca act gag tcc agc act gcc tcg 1764
 Ala Ala Tyr Ser Val Tyr Leu Asn Ser Thr Glu Ser Ser Thr Ala Ser
 275 280 285

 atc ctc ttc ggt gcg gtt gac caa agc aaa tat acc gga agt ctt ggc 1812
 Ile Leu Phe Gly Ala Val Asp Gln Ser Lys Tyr Thr Gly Ser Leu Gly
 290 295 300

 ttg ctt cct atc atc aac acg gct gct tcc tac ggt tac caa aag cct 1860
 Leu Leu Pro Ile Ile Asn Thr Ala Ala Ser Tyr Gly Tyr Gln Lys Pro
 305 310 315 320

 cta agg ctc caa atc acc ctg tct gcc att acg gtc agc gac tcc aga 1908
 Leu Arg Leu Gln Ile Thr Leu Ser Ala Ile Thr Val Ser Asp Ser Arg
 325 330 335

 gga cag caa gca agc att ggt tca gga gct gct gct gca ctt ctt gat 1956
 Gly Gln Gln Ala Ser Ile Gly Ser Gly Ala Ala Ala Leu Leu Asp
 340 345 350

Sequence Listing

acc gga acg act ttg acg tat gct cca agc gag att gtc gag aaa ctt			2004
Thr Gly Thr Thr Leu Thr Tyr Ala Pro Ser Glu Ile Val Glu Lys Leu			
355	360	365	
gct gaa acc cta ggc ttc gac tac agc agc tct gtc ggg gcc tac gtg			2052
Ala Glu Thr Leu Gly Phe Asp Tyr Ser Ser Val Gly Ala Tyr Val			
370	375	380	
gca aga tgc agg gac gtt gat agc tac gct gtc aac ttc gac ttc cag			2100
Ala Arg Cys Arg Asp Val Asp Ser Tyr Ala Val Asn Phe Asp Phe Gln			
385	390	395	400
ggt aaa gtg att gaa gct cct ttg agt tcc ttc ctg att gct ctg caa			2148
Gly Lys Val Ile Glu Ala Pro Leu Ser Ser Phe Leu Ile Ala Leu Gln			
405	410	415	
acc aac tcc gga gaa gtt tcc tcc tac tgc gca ttg ggt att ttc tcc			2196
Thr Asn Ser Gly Glu Val Ser Ser Tyr Cys Ala Leu Gly Ile Phe Ser			
420	425	430	
cct gga gac gaa tcc ttc acg ctc ggc gat act ttc ctg cga aac gcc			2244
Ser Gly Asp Glu Ser Phe Thr Leu Gly Asp Thr Phe Leu Arg Asn Ala			
435	440	445	
tac ttt gtg gct gac ctc gag gga tatcaa atc gct ata gct aac gtg			2292
Tyr Phe Val Ala Asp Leu Glu Gly Tyr Gln Ile Ala Ile Ala Asn Val			
450	455	460	
ac ctg aat cct gga gcc gag caa att gag gtc atc tca ggc aac tcc			2340
Asn Leu Asn Pro Gly Ala Glu Gln Ile Glu Val Ile Ser Gly Asn Ser			
65	470	475	480
tt cct tct gct tcg tcg gtt tcc gat tac tcc aat acc tgg ggc gcc			2388
Le Pro Ser Ala Ser Ser Val Ser Asp Tyr Ser Asn Thr Trp Gly Ala			
485	490	495	
ct gcc acc gct ttg gac act gac agg cct act act ctg gga tct gtg			2436
Ter Ala Thr Ala Leu Asp Thr Asp Arg Pro Thr Thr Leu Gly Ser Val			
500	505	510	

Sequence Listing

```

act gct gtg ggc gat gaa aga gtg acc tcg acc aag aag gtt tcg agt      2484
Thr Ala Val Gly Asp Glu Arg Val Thr Ser Thr Lys Lys Val Ser Ser
      515           520           525

gtg aag aca aac act tcg tcc ggg tcc ggg tcc act tcg gag tcg tct      2532
Val Lys Thr Ser Thr Ser Gly Ser Gly Ser Thr Ser Glu Ser Ser
      530           535           540

acg tcc agt tcg cat tcc aac aat ggc cca agg aca gta ggc ttt agt      2580
Thr Ser Ser His Ser Ser Asn Gly Pro Arg Thr Val Gly Phe Ser
      545           550           555           560

ttg tgt gcc gtt ttg tgc gca ttc ttg att tct ata cta gtt gtt tgc      2628
Leu Cys Ala Val Leu Cys Ala Phe Leu Ile Ser Ile Leu Val Val Cys
      565           570           575

ta gatctgaagt tctaaaggggc ttttagtcttc atttatgatt tttttttatt      2680

tggaccgcct cgaattgttt ttccgacggg tctactttaa agctgcaaga tctcgtttag      2740

cgtcgtttat ttctcgttcg ttttagtgaca aaaaaacaga aaaaaaaact ataaaaagcg      2800

gtatataacc tttatatttt gataaacatg agcagcgaaa ttaagctagc accaaaggat      2860

tacgagaagg acaaggagtt cgccaaaggct ctgcatggca aggacgccc gagcgctaca      2920

ggaatgagtg cttgggtgaa gaaggacaag gaagctaaa aagtgcgcgt ggaaggatat      2980

ttcaagcact gggacgggaa aaccgacgag gagactgaaa agtgcgagact cgaggactac      3040

tcgacgctca ccaagcacta ctacaacctg gtgacggatt tctacgagta tggatgggaa      3100

tcctcggtcc actttccag atactacaag ggagagccat ttagacaagc t      3151

<210> 2
<211> 576
<212> PRT
<213> Hansenula polymorpha

```

Sequence Listing

<400> 2

Met Lys Val Ala Thr Leu Phe Phe Leu Ala Ser Ser Val Cys Val Leu

1 5 10 15

Gly Asp Pro Gln Phe Val Lys Leu Glu Ala Ser Val Leu Arg Gly Ser

20 25 30

Thr Tyr Lys Asp Ser Gln Lys Gly Ala Lys Pro Phe Met Leu Glu Lys

35 40 45

Arg Ala Asp Asp Gly Ser Val Thr Met Glu Leu Gln Asn Ala Gln Ser

50 55 60

Phe Tyr Gln Val Glu Ile Glu Ile Gly Ser Asp Lys Gln Lys Val Gly

65 70 75 80

Val Leu Ile Asp Thr Gly Ser Ser Asp Leu Trp Val Met Asn Ser Asn

85 90 95

Asn Ser Tyr Cys Ser Ser Ser Ser Thr Lys Lys Leu Lys Arg Asp Gly

100 105 110

Pro Ala Asp Ala Leu Gln Lys Gly Arg Asp Leu Ser Asp Leu Tyr Asn

115 120 125

Phe Asn Ser Pro Asn Glu Asp Asn Asn Ala Lys Gly Phe Leu Gly Gly

130 135 140

Trp Gly Asp Leu Thr Thr Val Glu Thr Ala Thr Gln Asp Glu Thr Gln

145 150 155 160

Thr Ala Leu Ala Ala Gln Ala Thr Val Asp Cys Ser Leu Tyr Gly Thr

165 170 175

Phe Asn Pro Ser Thr Ser Asn Ser Phe His Asn Asn Gly Thr Thr Phe

180 185 190

Glu Ile Ser Tyr Ala Asp Arg Thr Phe Ala Arg Gly Thr Trp Gly Tyr

195 200 205

Sequence Listing

Asp Asp Val Thr Phe Asn Gly Val Thr Val Asn Asp Leu Ser Leu Ala
210 215 220

Val Ala Asp Glu Thr Asp Ser Ser Thr Gly Val Phe Gly Ile Gly Leu
225 230 235 240

Arg Glu Leu Glu Thr Thr Tyr Ser Gly Gly Pro Gln His Tyr Ile
245 250 255

Tyr Asp Asn Leu Pro Phe Lys Met Val Asp Gln Gly Leu Ile Asn Arg
260 265 270

Ala Ala Tyr Ser Val Tyr Leu Asn Ser Thr Glu Ser Ser Thr Ala Ser
275 280 285

Ile Leu Phe Gly Ala Val Asp Gln Ser Lys Tyr Thr Gly Ser Leu Gly
290 295 300

Leu Leu Pro Ile Ile Asn Thr Ala Ala Ser Tyr Gly Tyr Gln Lys Pro
305 310 315 320

Leu Arg Leu Gln Ile Thr Leu Ser Ala Ile Thr Val Ser Asp Ser Arg
325 330 335

Gly Gln Gln Ala Ser Ile Gly Ser Gly Ala Ala Ala Leu Leu Asp
340 345 350

Thr Gly Thr Thr Leu Thr Tyr Ala Pro Ser Glu Ile Val Glu Lys Leu
355 360 365

Ala Glu Thr Leu Gly Phe Asp Tyr Ser Ser Ser Val Gly Ala Tyr Val
370 375 380

Ala Arg Cys Arg Asp Val Asp Ser Tyr Ala Val Asn Phe Asp Phe Gln
385 390 395 400

Gly Lys Val Ile Glu Ala Pro Leu Ser Ser Phe Leu Ile Ala Leu Gln
405 410 415

Sequence Listing

Thr Asn Ser Gly Glu Val Ser Ser Tyr Cys Ala Leu Gly Ile Phe Ser
420 425 430

Ser Gly Asp Glu Ser Phe Thr Leu Gly Asp Thr Phe Leu Arg Asn Ala
435 440 445

Tyr Phe Val Ala Asp Leu Glu Gly Tyr Gln Ile Ala Ile Ala Asn Val
450 455 460

Asn Leu Asn Pro Gly Ala Glu Gln Ile Glu Val Ile Ser Gly Asn Ser
465 470 475 480

Ile Pro Ser Ala Ser Ser Val Ser Asp Tyr Ser Asn Thr Trp Gly Ala
485 490 495

Ser Ala Thr Ala Leu Asp Thr Asp Arg Pro Thr Thr Leu Gly Ser Val
500 505 510

Thr Ala Val Gly Asp Glu Arg Val Thr Ser Thr Lys Lys Val Ser Ser
515 520 525

Val Lys Thr Ser Thr Ser Ser Gly Ser Gly Ser Thr Ser Glu Ser Ser
530 535 540

Thr Ser Ser Ser His Ser Ser Asn Gly Pro Arg Thr Val Gly Phe Ser
545 550 555 560

Leu Cys Ala Val Leu Cys Ala Phe Leu Ile Ser Ile Leu Val Val Cys
565 570 575

<210> 3
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

Sequence Listing

<400> 3

gaagtgcagc agcagtcctt gaacc

25

<210> 4

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 4

ggctgatgac ggctcggtca cgatgg

26

<210> 5

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 5

ggacacgcaa gaggtgtctg

20

<210> 6

<211> 40

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

Sequence Listing

<400> 6
agctcgctac ccggggatcc gcaactttca ttgtgtcaac 40

<210> 7
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 7
gcacatcccc ctttcgccag cctttcggt gcgggtgacc 40

<210> 8
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 8
gctcggtcagg 20

<210> 9
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

Sequence Listing

<400> 9
ggatccccgg gtaccgagct

20

<210> 10
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 10
caccggtagc taatgatccc

20

<210> 11
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 11
cgaacatcca agtgggccga

20

<210> 12
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 12

Sequence Listing

ctggcgaaag gggatgtgc

20

<210> 13
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 13
gaattcatga agtgggtaac cttt

24

<210> 14
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 14
taaggctaag gcagcttgac

20

<210> 15
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 15
caagctgcct taggcttatg cagctgtcc ccgggtg

36

Sequence Listing

<210> 16
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 16
actatgtgatt tatgggtccct cgatg

25